انجام پروژه داده کاوی

آکادمی داده کاوی مجموعه‌ای از برترین متخصصان مجرب است که در زمینه انجام پروژه‌های داده کاوی در تمامی سطوح فعالیت می‌کند. این مجموعه با سه سال تجربه موفق در زمینه علاوه بر انجام پروژه داده کاوی شرکتی و صنعتی ، در زمینه آموزش انجام پروژه دانشجویی داده کاوی و آموزش انجام پایان نامه به صورت تخصصی در تمامی رشته‌ها فعالیت می‌کند . ما در آکادمی برآنیم علم داده کاوی را به طور گسترده در امور بازاریابی ، فروش ، پزشکی و ... در ایران را گسترده و کاربردی کنیم .

داده کاوی در واقع علم کشف دانش از حجم وسیعی از داده می‌باشد. مثالی که نزدیک‌ترین شباهت به علم داده کاوی را داشته باشد میتوان معدن کاری را نام برد . کشف طلا از حجم گسترده‌ای صخره‌ها و کوه‌ها ، برای کشف دانش از طریق داده کاوی از الگوریتم‌های داده کاوی و نرم افزارهای داده کاوی استفاده می‌شود.

خدمات ما در زمینه انجام پروژه با نرم افزار داده کاوی

- انجام پروژه داده کاوی با رپیدماینر
- انجام پروژه داده کاوی با پایتون
- انجام پروژه داده کاوی با متلب
- انجام پروژه داده کاوی با وکا
- انجام پروژه داده کاوی با spss
- انجام پروژه داده کاوی با R
- انجام پروژه داده کاوی با کلمنتاین
- انجام پروژه داده کاوی با orange

آکادمی داده کاوی به همراه داشتن متخصصانی بسیار مجرب و فارغ التحصیل از دانشگاه‌های برتر کشور در زمینه کاری خود دارای مقاله‌های متعدد ISI می‌باشد.آکادمی داده کاوی در زمینه آموزش خدمات زیر را ارائه میدهد .

- آموزش انجام پایان نامه در تمامی رشته ها
- آموزش انجام پروژه‌های دانشجویی داده کاوی
- آموزش پروپوزال نویسی
- آموزش نوشتن مقاله ISI از پایان نامه
- کمک در انتخاب موضوع پایان نامه
- انجام پروژه یادگیری ماشین
- پروژه در مورد داده کاوی
- انجام پروژه یادگیری عمیق
- انجام پروژه بیگ دیتا
- انجام پروژه شبکه عصبی
- انجام پروژه هوش مصنوعی
- انجام پروژه متن کاوی

- انجام تمرین داده کاوی

مشاوره_هر زمانی به هر شکلی که برای شما آسانتر است:

داده‌کاوی

تعریف داده کاوی

داده کاوی یک فرآیند محاسباتی است که در واقع الگو یا الگوهایی را در مجموعه از داده های عظیم کشف میکند . در تمامی تعریف های مرتبط به داده کاوی کلمه کشف کردن را میتوان پیدا کرد . داده کاوی شاخه ای از علوم کامپیوتر است که در واقع ترکیبی از تکنیکهای آماری ، علوم اطلاعات ، یادگیری ماشینی و نظریه پایگاه داده است .

تاریخچه داده کاوی

اصطلاح Data Fishing یا Data Dredging به معنی صید داده بوجود آمد . این اصطلاح به این معنی است که درون حجم عظیمی از داده ها بدون در نظر گرفتن هیچگونه پیش فرضی ، هر گونه ارتباطی را مورد بررسی قرار دهیم. در سال ۱۹۸۹ اصطلاح کشف دانش در پایگاه داده مطرح شد و در سال ۱۹۹۰ اصطلاح داده کاوی بوجود آمد و در همین سال با استفاده از داده کاوی ، خرده فروش ها و بازارهای مالی به تجزیه تحلیل داده ها و پیش بینی نوسانات در نرخ بهره و افزایش مشتری پرداختند .

تفاوت داده کاوی با آمار

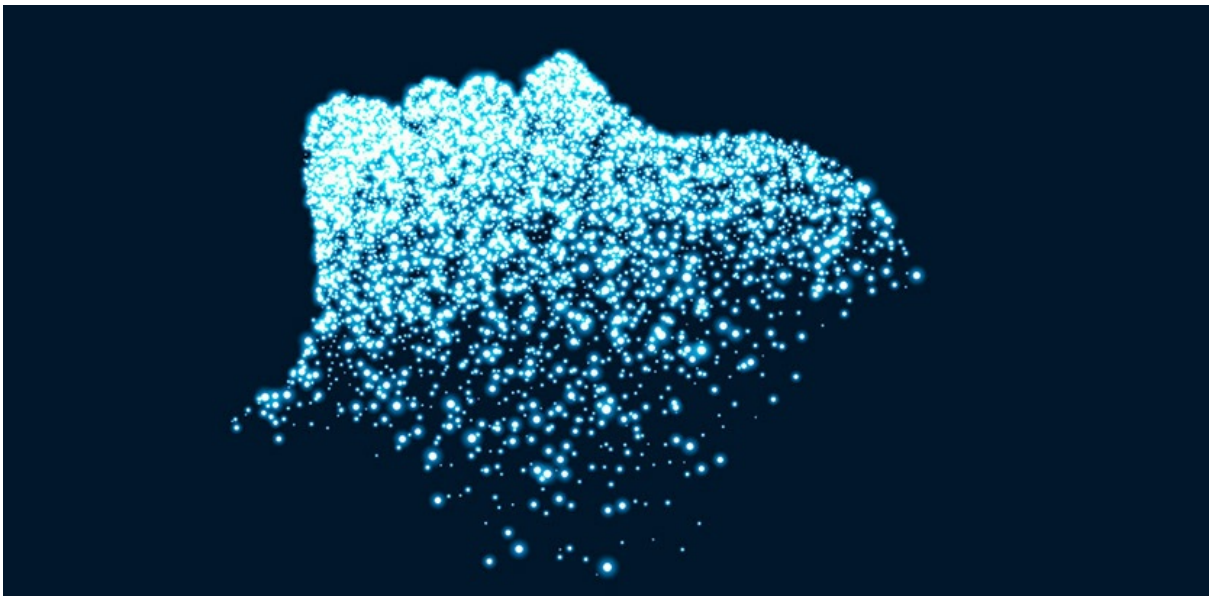
تحلیل های آماری در کل بنظر بی شباهت به تحلیل های داده کاوی نیستند . اما حالب ترین تفاوت این دو با هم در این است که در آمار پیش فرض هایی وجود دارد و آمار گر با تحلیل خود به اثبات یا رد آن فرضیه میپردازد اما در داده کاوی گویی داده کاو خود نیز نمیداند دقیقا دنبال چه چیزی هست ! او به دنبال ارتباط های پنهان و کشف نشده است . داده کاوی پل ارتباطی میان علم آمار ، علم کامپیوتر ، هوش مصنوعی ، الگوشناسی ، فراگیری ماشین داده می باشد.

چه چیزی اساسا باعث پیدایش داده کاوی شد ؟

میتوان گفت اصلی ترین دلیل و توجه به این دانش این بود که حجم وسیعی از داده ها موجود بود و شدیدا این نیاز بوجود آمده بود که باید از این داده ها اطلاعات و دانش مفیدی استخراج کرد و از آن برای کنترل ، تحلیل و پیش بینی استفاده کرد بنابراین دلایل کلی را میتوان در سه مورد نوشت : حجم داده‌ها (Data) با سرعت زیادی در حال رشد است. اطلاعات (Information) ما در مورد این داده‌ها کم است. دانش (Knowledge) ما نسبت به این اطلاعات صفر است. اما پاسخ تمامی این مشکلات تنها و تنها داده کاوی است . با داده کاوی میتوان حجم عظیم داده ها را اصطلاحا کاوید و دانش را از دل آن بیرون کشید .

ویژگی های اصلی داده کاوی

- کشف اتوماتیک الگوها در انجام پروژه های داده کاوی
- پیشبینی احتمالی نتایج و خروجی ها
- تولید اطلاعات اجرایی و مفید
- متمرکز بر داده های بزرگ
- مزایای داده کاوی
- ایجاد روابط بصورت اتوماتیک
- استفاده از داده های متنوع
- دینامیک بودن
- عدم نیاز به داده های صحیح
- ساخت مدل های واقعی
- آنالیز کردن داده های واقعی
- دوری از اشکالات احتمالی نمونه گیری



اجزاي اصلي سيستم داده كاوی

پایگاه داده: مجموعه‌اي از پايگاه داده ها، انباره داده، صفحه گسترده، يا ديگر انواع مخازن اطلاعات، پاكسازي داده‌ها و تكنيك‌هاي يكپارچه سازي روي اين داده‌ها انجام مي‌شود. سرويس دهنده پايگاه داده: كه مسئول بازيابي داده‌هاي مرتبط براساس نوع درخواست داده كاوي كاربر مي‌باشد.

پایگاه دانش: این پایگاه از دانش زمینه تشکیل شده تا به جستجو کمك کند یا براي ارزیابي الگوهاي یافته شده از آن استفاده مي‌شود.

موتور داده کاوي: این موتور جزء اصلي سیستم داده کاوي است و به طور ایده آل شامل مجموعه‌اي از پیمانه‌ها نظیر توصیف، تداعي، کلاس بندي، آنالیز خوشه‌ها و آنالیز تکامل و انحراف است.

پیمانه ارزیابي الگو: این جزء معیارهاي جذابیت را به کار مي‌بندد و با پیمانه داده کاوي تعامل مي‌کند، بدین صورت که تمرکز آن بر جستجو بین الگوهاي جذاب مي‌باشد، و از یك حد آستانه جذابیت استفاده مي‌کند تا الگوهاي کشف شده را ارزیابي کند.

واسط گرافیکي کاربر: این پیمانه بین کاربر و سیستم داده کاوي ارتباط برقرار مي‌کند، به کاربر اجازه مي‌دهد تا با سیستم داده کاوي از طریق پرس و جو ارتباط برقرار کند. این جزء به کاربر اجازه مي‌دهد تا شماي پایگاه داده یا انباره داده را مرور کرده، الگوهاي یافته شده را ارزیابي کرده و الگوها را در فرم‌هاي بصري گوناگون، بازنمایي کند.

کاربردهای داده کاوی

علم داده کاوی جزو ده علم برتری است که تحولی در عصر تکنولوژی ایجاد نموده است . در هر فضایی که داده ای تولید میشود میتوان گفت داده کاوی میتواند کاربرد داشته باشد . از جمله امور تجاری و مالی، امور پزشکی، زیست پزشکی، تجزیه و تحلیلهای مربوط به DNA، ورزش و سرگرمی، کشف ناهنجاریها و اسناد جعلی، ارتباطات از راه دور، کتابداری و اطلاع‌رسانی ... امروزه استفاده از داده کاوی در شرکتها به امری حیاتی تبدیل شده است . تمامی شرکتهایی که پارامتر مشتریان در کانون توجه است مانند فروشگاه‌ها، شرکتهای مالی، ارتباطاتی، بازاریابی و غیره استفاده از داده کاوی برای کشف اطلاعات ضروری است . این اطلاعات حاصل از داده کاوی به شرکتها کمک میکند تا ارتباطهای عوامل داخلی قیمت، محل قرارگیری محصولات و مهارت کارمندان را با عوامل خارجی از جمله: وضعیت اقتصادی، رقابت در بازار و محل جغرافیایی مشتریان کشف نمایند. انجام پروژه های داده کاوی پیش‌بینی وضع آینده بازار، گرایش مشتریان و شناخت سلیقه‌های عمومی آنها را برای شرکتها ممکن می‌سازد .

بگذارید چند مثال در زمینه های مختلف بزنیم و هر کدام را مقداری توضیحی بدهیم .

در زمینه صنعت

یکی از مثالهای معروف استفاده از داده کاوی در صنعت متعلق به شرکت فولادسازی پوهانگ است . که با استفاده از الگوریتمهای داده کاوی حدود ۱۵ درصد میزان مصرف انرژی را کاهش دادند که نتیجه این کاهش ۳/۱ میلیون دلار صرفه جویی بوده واین نتیجه قطعا باعث کاهش قیمت محصولات در نتیجه افزایش فروش و رضایتمندی بیشتر مشتریان شده است .

در هتل‌داری

در صنعت هتلداری استفاده از داده کاوی میتواند فرق برنده و بازنده را مشخص کند . یکی از هتلهای مشهوری که در لاس وگاس است ؛ برای افزایش رضایتمندی مسافران از الگوریتمهای داده کاوی بهره برد . این هتل بصورت پرسشنامه داده هایی جمع نمود و توانست عواملی که باعث میشود مسافران دوباره به هتل بازگردند را پیدا کرده و با طبقه‌بندی مسافران، مسافران وفادار به هتل را پیدا کنند.

در مدیریت ریسک

در بانکی از بامکهای کانادا مسئله مهم تقلب در حسابها و چگونگی و میزان برگشت وامهای داده شده توسط بانک با استفاده از الگوریتم های داده کاوی را حل نمودند . حوزه‌های داده‌کاوی در سه حوزه مستقل به‌کار می‌رود و در آنها ریشه دوانده است: ۱ر آمار کلاسیک و الگوهای آماری ۲ر هوش مصنوعی ۳ر یادگیری خودکار و شبکه‌های عصبی

مراحل داده کاوی

مرحله اول: تشکیل انبار داده. با توجه به عنوان، این مرحله برای تشکیل محیطی پیوسته و یک پارچه جهت انجام مراحل بعدی و داده کاوی در آن، انجام می گیرد. در حالت کلی انبار داده مجموعه پیوسته و طبقه بندی شده است که دائما در حال تغییر بوده و دینامیک است که برای کاوش آماده می شود.

مر حله دوم: انتخاب داده ها. در این مرحله برای کم کردن هزینه های عملیات داده کاوی، داده هایی از پایگاه داده انتخاب می شوند. که مورد مطالعه هستند و هدف داده کاوی دادن نتایجی در مورد آن هاست.

مرحله سوم: تبدیل داده ها. مشخص است برای انجام عملیات داده کاوی لزوما باید تبدیلات خاصی روی داده ها انجام گیرد ممکن است این تبدیلات خیلی راحت و مختصر مثل تبدیل byte به integer باشد یا خیلی پیچیده و زمان بر و با هزینه های بالا مثل تعریف صفات جدید و یا تبدیل و استخراج داده ها از مقادیر رشته ای و ... باشد.

مرحله چهارم: کاوش در داده ها. در این مرحله است که داده کاوی انجام می شود. در این مرحله با استفاده از تکنیک های داده کاوی داده ها مورد کاوش قرار گرفته، دانش نهفته در آن ها استخراج شده و الگو سازی صورت می گیرد.

مرحله پنجم: تفسیر نتیجه در این مرحله نتایج و الگوهای ارائه شده توسط ابزار داده کاوی مورد بررسی قرار گرفته و نتایج مفید معین می شود. طرز کار ابزار داده کاوی این گونه است که ابزار به دنبال اثبات این است که وجود چیزی به معنای وجود چیز دیگری است و سعی می کند در درجه اول از توالی ارتباطات برای کشف یک الگو بهره بگیرد و در نهایت اطلاعات بدست آمده را دسته بندی کند تا به الگوی خاصی برسد که بتواند آن را براساس فاکتورهای داخلی به مخاطبش ارائه دهد.

CRISP-DM (انجام پروژه داده کاوی)

برای انجام پروژه داده کاوی یکی از متداولترین روش ها ، CRISP-DM (Cross-Industry Standard Process for Data Mining) نام دارد. در میانه دهه ۱۹۹۰ میلادی نخستین بار این استاندارد توسط گروهی از شرکت‌های اروپایی به‌عنوان روشی برای انجام پروژه داده کاوی به جهان ارائه شد و از آن زمان تاکنون این روش در سرتاسر جهان توسط تحلیلگران مورد استفاده قرار میگیرد. فرآیند کریسپ دارای شش مرحله است که از درک نیازهای اصلی یک کسب‌وکار آغاز می‌شود و در نهایت به ارائه راهکاری برای آن نیازی که از ابتدا مطرح شده است ختم می‌شود. بنظر می رسد مراحل این فرآیند که ذکر شده است به دنبال یکدیگر می‌آیند اما در عمل دقیقا به این شکل نیست و در واقع رفت‌وبرگشت‌های زیادی بین مراحل مختلف فرآیند کریسپ وجود دارد. تحلیلگران به‌خوبی به این مطلب واقف هستند که برای کار با داده نیازمند سعی و خطا و آزمایش کردن بسیار است.

اگر متن های قدیمی و کلاسیک داده کاوی را مطالعه کرده باشید، در آن چهار وظیفه ی اساسی برای داده کاوی برشمرده شده است

- دسته بندی (Classification)
- خوشه بندی (Clustering)
- تحلیل قواعد انجمنی (Association Rules)
- مصور سازی (Visualization)

که البته کمابیش نقش های دیگری نیز به ان اضافه می کنند ولی در کل همین است که همین است. خب این یک تقسیم بندی. در تقسیم بندی دیگر، تمام الگوریتم های داده کاوی و یادگیری ماشین به سه دسته کلی:

- یادگیری با ناظر (Supervised Learning)
- یادگیری بدون ناظر (Unsupervised Learning)
- یادگیری نیمه ناظر (Semi-Supervised Learning)

هر کدام از این تقسیم بندی ها الگوریتم ها و دنیای خودش را دارد ولی به طور کلی، یادگیری با ناظر با استفاده از برچسب یا label داده ها، برچسب داده های مشاهده نشده را تشخیص می دهد ولی در یادگیری بدون ناظر معمولا برچسب داده ها موجود نیست. یادگیری نیمه ناظر هم چیزی بین این دوتاست که معمولا حالت استاد-شاگردی در آن شبیه سازی می شود و استاد تنها در بعضی مواقع خاص تقلب هایی

به شاگردش که الگوریتم باشد، می رساند. هر کدام از این سه حالت کلی، دنیای خاص خودشان را دارند. به طور مثال خود یادگیری باناظر در یک دسته بندی به زیربخش های زیر می تواند تفکیک شود: (بر مبنای کتاب Machine Learning With R)

یادگیری تنبل (Lazy Learning): که با استفاده از ذخیره سازی داده ها، دسته بندی را انجام می دهند. مثال بارزش الگوریتم نزدیک ترین همسایگی (KNN) است.

یادگیری احتمالی (Probabilistic Learning): با استفاده از تکنیک ها و روش های آماری نظیر تئوری بیزین، کار دسته بندی را انجام می دهند.

یادگیری تقسیم و غلبه (Divide and Conquer Learning): با استفاده از استراتژی معروف انگلیس در سال های دور و شاید الان، یعنی تقسیم و غلبه، دسته بندی صورت می گیرد. مثال معروفش درخت تصمیم است.

پیش بینی مقدار عددی (Regression Methods): در اینجا برچسب خود مقدار متغیر پاسخ است که با استفاده از الگوریتم های مختلف، پیش بینی می شود. مثال معروفش همان رگرسیون خطی است.

روش های جعبه سیاهی (Black-box Methods): که نحوه ی عملکرد دقیق در میانه های راه اجرای الگوریتم ها قابل تفکیک نیست که عمدتا ناشی از پیچیدگی ذات متغیرها و تعداد زیاد آنهاست. اصلی ترین روش شبکه عصبی و خانواده ی پرجمعیت آن است. در یادگیری ماشین و داده کاوی هیچ وحی منزل ای وجود ندارد لذا این تقسیم بندی یک تقسیم بندی پیشنهادی است. بنابراین با این فرض می توان حوزه ی کلی یادگیری بدون ناظر را نیز به زیربخش های زیر تقسیم کرد: (بر مبنای کتاب Data Mining Concepts and Techniques)

- خوشه بندی افرازی (Partitioning Methods)
- خوشه بندی سلسه مراتبی (Hierarchical Methods)
- خوشه بندی چگالی محور (Density-Based Methods)
- خوشه بندی شبکه ای (Grid-Based Methods)

- الگوریتم C4.5
- الگوریتم k-means
- لگوریتم Support vector machines

- الگوریتم Apriori
- الگوریتم EM
- الگوریتم PageRank
- الگوریتم AdaBoost
- الگوریتم kNN
- الگوریتم Naive Bayes
- الگوریتم CART

الگوریتم های داده کاوی

- الگوریتم وابستگی (Association algorithm)
- الگوریتم خوشه‌بندی (Clustering algorithm)
- الگوریتم درخت تصمیم (Decision Trees algorithm)
- الگوریتم رگرسیون خطی (Linear Regression algorithm)
- الگوریتم بیز (Naive Bayes Algorithm)
- الگوریتم شبکه‌های عصبی (Neural Network Algorithm)
- الگوریتم رگرسیون منطقی یا لجستیک (Logistic regression algorithm)
- الگوریتم خوشه‌بندی زنجیره‌ای (Sequence Clustering algorithm)

نرم افزارهای داده کاوی

انجام پروژه داده کاوی با نرم افزار RapidMiner: این نرم‌افزار که به گفته سازندگان آن تلاش بر ین کرده است که به صورت یکپارچه عملیاتِ مختلفِ حوزه‌ی علوم‌داده را تجمیع کند و به دانشمندانِ علوم داده اجازه دهد به سرعت مدل‌های مورد نیاز برای عملیات داده‌کاوی را شناسایی کنند.

انجام پروژه داده کاوی با نرم افزار Weka: نرم‌افزار وکا(weka) مجموعه‌ای از الگوریتم‌های مختلف جهت عملیات داده‌کاوی را در اختیار متخصصان و دانشمندانِ علوم داده می‌گذارد. کار با این نرم‌افزار بسیار ساده است و در اینجا کتابی جهت آموزش نرم افزار weka توسط خود سایت سازنده قرار داده شده است.

انجام پروژه داده کاوی با نرم افزار Orange: یکی از نرم‌افزارهای بسیار ساده و لذت بخش جهت انواع عملیات داده‌کاوی نرم افزاری Orange است. این نرم افزار به خاطرِ سادگی و واسطِ کاربریِ ساده‌ی آن می‌تواند موردِ استفاده‌ی بسیاری از متخصصانِ حوزه‌ی علوم‌داده باشد. حتی دوستانی که به تازگی به دنبالِ یادگیریِ علوم داده هستند، می‌توانند از این نرم افزار استفاده کنند.

انجام پروژه داده کاوی با نرم افزار Neural Designer: مخصوص طراحی شبکه های عصبی اگر با شبکه‌های عصبی آشنا باشید می‌دانید که طراحیِ این گونه شبکه‌ها معمولاً کارِ وقت‌گیری است و نیاز به دقتِ بالایی دارد. با استفاده از نرم افزار Neural Designer به راحتی می‌توانید شبکه‌های عصبیِ مخصوصِ خود را طراحی کنید و مدل‌های مختلفِ داده را توسط آن ها آزمایش کنید.

انجام پروژه داده کاوی با پایتون: انجام پروژه های پایتون با فیلدهایی مانند یادگیری ماشین (یادگیری با نظر و بدون ناظر ، خوشه بندی ، دسته بندی) همچنین انجام پروژه داده کاوی با پایتون (پیش بینی در دیتا ، کشف تقلب و ...) همچنین انجام پروژه در شناسایی الگو و پردازش تصویر

انجام پروژه های داده کاوی با متلب: این نرم افزار یکی دیگر از نرم افزارهایی است که در امر انجام پروژه داده کاوی استفاده میشود. برای مثال امروزه دانشجویان مهندسی در پایان نامه خود از این نرم افزار سود میبرند .